BỘ GIÁO DỤC VÀ ĐÀO TẠO



**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HỒ CHÍ MINH**

**ĐỒ ÁN CƠ SỞ**

**TÌM HIỂU TỔNG QUAN VỀ CÁC KỸ THUẬT NHẬN DẠNG KHUÔN MẶT VÀ ỨNG DỤNG THỰC NGHIỆM**

Ngành : **CÔNG NGHỆ THÔNG TIN**

Chuyên ngành: **KHOA HỌC DỮ LIỆU**

Giảng viên hướng dẫn : ThS. Lê Nhật Tùng

Sinh viên thực hiện :

+ Nguyễn Đức Trường MSSV: 2286400868 Lớp: 22DKHA1

+ Trần Tuấn Đạt MSSV: 2286400007 Lớp: 22DKHA1

+ Lê Trương Duy Khôi MSSV: 2286400012 Lớp: 22DKHA1

TP. Hồ Chí Minh, 2025

BỘ GIÁO DỤC VÀ ĐÀO TẠO



**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HỒ CHÍ MINH**

**ĐỒ ÁN CƠ SỞ**

**TÌM HIỂU TỔNG QUAN VỀ CÁC KỸ THUẬT NHẬN DẠNG KHUÔN MẶT VÀ ỨNG DỤNG THỰC NGHIỆM**

Ngành : **CÔNG NGHỆ THÔNG TIN**

Chuyên ngành: **KHOA HỌC DỮ LIỆU**

Giảng viên hướng dẫn : ThS. Lê Nhật Tùng

Sinh viên thực hiện :

+ Nguyễn Đức Trường MSSV: 2286400868 Lớp: 22DKHA1

+ Trần Tuấn Đạt MSSV: 2286400007 Lớp: 22DKHA1

+ Lê Trương Duy Khôi MSSV: 2286400012 Lớp: 22DKHA1

TP. Hồ Chí Minh, 2025

**NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN**

…………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………….…………………………………..……………………………………………….………………………………………………………………...…………………….…………………………………………………………………………………….…………………………………………………………………………………….………………………………..…………………………………………………….……………………………………………………………………………………………………

|  |
| --- |
| TPHCM, Ngày … tháng … năm 2024 |
| **Giáo viên hướng dẫn** |
| (Ký tên, đóng dấu) |

## LỜI CAM ĐOAN

Chúng tôi, gồm ba thành viên là Trần Tuấn Đạt, Lê Trương Duy Khôi, Nguyễn Đức Trường.

Xin cam đoan rằng mọi thông tin được trình bày trong báo cáo này đều chính xác và đầy đủ nhất theo sự hiểu biết của chúng tôi. Toàn bộ nội dung của báo cáo đều được trình bày dựa trên quan điểm, kiến thức cá nhân và tích lũy của từng thành viên trong nhóm và được chọn lọc từ nhiều nguồn tài liệu có đính kèm chi tiết và hợp lệ. Chúng tôi cũng cam đoan rằng bài báo cáo này không sao chép từ bất kỳ nguồn nào mà không được chỉ rõ trong văn bản hoặc được đánh dấu trong phần tài liệu tham khảo. Chúng tôi hoàn toàn chịu trách nhiệm về nội dung của báo cáo và sẵn sàng bổ sung hoặc sửa lỗi nếu có yêu cầu, nhằm đảm bảo tính trung thực và trách nhiệm đối với báo cáo này.

Xin chân thành cảm ơn!

MỤC LỤC

[LỜI CAM ĐOAN ii](#_Toc196469182)

[DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT iv](#_Toc196469183)

[DANH MỤC CÁC BẢNG, HÌNH VẼ, ĐỒ THỊ v](#_Toc196469184)

[LỜI MỞ ĐẦU 1](#_Toc196469185)

[CHƯƠNG 1 TỔNG QUAN 2](#_Toc196469186)

[1.1 Giới thiệu đề tài 2](#_Toc196469187)

[1.2 Nhiệm vụ của đề tài 2](#_Toc196469188)

[1.2.1 Tính cấp thiết của đề tài 3](#_Toc196469189)

[1.2.2 Ý nghĩa khoa học và thực tiễn của đề tài 4](#_Toc196469190)

[1.3 Mục tiêu nghiên cứu 5](#_Toc196469191)

[1.3.1 Mục tiêu tổng quát 5](#_Toc196469192)

[1.3.2 Mục tiêu cụ thể 5](#_Toc196469193)

[1.4 Đối tượng và phạm vi nghiên cứu 5](#_Toc196469194)

[1.4.1 Đối tượng nghiên cứu 5](#_Toc196469195)

[1.4.2 Phạm vi nghiên cứu 5](#_Toc196469196)

[1.5 Phương pháp nghiên cứu 6](#_Toc196469197)

[1.5.1 Phương pháp nghiên cứu sơ bộ 6](#_Toc196469198)

[1.5.2 Phương pháp nghiên cứu tài liệu 6](#_Toc196469199)

[1.5.3 Phương pháp nghiên cứu thống kê 6](#_Toc196469200)

[1.5.4 Phương pháp thực nghiệm 7](#_Toc196469201)

[1.5.5 Phương pháp đánh giá 7](#_Toc196469202)

[1.6 Những đóng góp nghiên cứu của đề tài 7](#_Toc196469203)

[1.6.1 Đóng góp về mặt lý thuyết 7](#_Toc196469204)

[1.6.2 Đóng góp trong thực tiễn 8](#_Toc196469205)

[CHƯƠNG 2 CƠ SỞ LÝ THUYẾT 9](#_Toc196469206)

[2.1 Giới thiệu tổng quan về thị giác máy tính (computer vision) 9](#_Toc196469207)

[2.1.1 Định nghĩa và vai trò của nhận dạng khuôn mặt 9](#_Toc196469208)

[2.1.2 Ưu điểm và hạn chế 9](#_Toc196469209)

[2.2. Deep Learning 10](#_Toc196469210)

[2.3. OpenCV (Open Source Computer Vision Library) 11](#_Toc196469211)

[2.4 Các phương pháp, kỹ thuật tiền xử lý ảnh. 12](#_Toc196469212)

[**2.4.1 Chuyển sang ảnh xám (Grayscale Conversion)** 12](#_Toc196469213)

[**2.4.2** **Cân bằng Histogram (Histogram Equalization)** 13](#_Toc196469214)

[**2.4.3** **Làm mịn ảnh (Image Smoothing-blurring)** 14](#_Toc196469215)

[**2.4.4** **Biến đổi hình học (Affine / Perspective Transformations)** 15](#_Toc196469216)

[2.4.4.1.1Biến đổi hình học (Affine Transformation) 15](#_Toc196469217)

[2.4.4.2 Biến đổi phối cảnh (Perspective Transformation) 15](#_Toc196469218)

[**2.4.5 Data Augmentation (Phóng đại dữ liệu)** 16](#_Toc196469219)

[**2.4.6 GAN-based Enhancement (Generative Adversarial Network)** 17](#_Toc196469220)

[2.5 Phát hiện đặc trưng và mô tả (Feature Detection & Description) 18](#_Toc196469221)

[**2.5.1 Edge Detection**: 18](#_Toc196469222)

[**2.5.2 Blob Detection:** 19](#_Toc196469223)

[**2.5.4 Learned Features** 19](#_Toc196469224)

[2.5.4.1 ResNet 20](#_Toc196469225)

[2.5.4.2 EfficientNet: 20](#_Toc196469226)

[**2.5.5 Vision Transformers (ViT)** 21](#_Toc196469227)

[2.6 Phát hiện đối tượng (Object Detection) 22](#_Toc196469228)

[**2.6.1** **R-CNN** 22](#_Toc196469229)

[**2.6.2 YOLO (You Only Look Once)** 22](#_Toc196469230)

[2.7 Nhận dạng và phân loại đối tượng (Object Classification & Recognition) 23](#_Toc196469231)

[**2.7.1 HOG Kết Hợp SVM** 23](#_Toc196469232)

[*2.7.1.1* *HOG - Histogram of Oriented Gradients* 23](#_Toc196469233)

[2.7.1.2 SVM - Support Vector Machine 24](#_Toc196469234)

[**2.7.2 Template Matching** 25](#_Toc196469235)

[**2.7.3 Transformer-based: Vision Transformer (ViT), Swin Transformer** 25](#_Toc196469236)

[2.7.3.1Swin Transformer 25](#_Toc196469237)

## DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

## DANH MỤC CÁC HÌNH ẢNH

## DANH MỤC CÁC BẢNG, HÌNH VẼ, ĐỒ THỊ

## LỜI MỞ ĐẦU

Trong cuộc sống hiện đại, công nghệ ngày càng phát triển và được ứng dụng rộng rãi vào nhiều lĩnh vực khác nhau và nó đóng vai trò quan trọng trong hầu hết các lĩnh vực của đời sống, từ giáo dục, y tế cho đến giao thông và sản xuất. Bên cạnh những công nghệ quen thuộc như trí tuệ nhân tạo hay học máy Một trong những lĩnh vực đang được quan tâm nhiều hiện nay là thị giác máy tính, nơi mà máy móc có thể nhìn và hiểu hình ảnh giống như con người.

Và với mong muốn tìm hiểu sâu hơn về lĩnh vực này, nhóm chúng em đã chọn đề tài **“**Tìm Hiểu Tổng Quan Về Các Kỹ Thuật Nhận Dạng Khuôn Mặt Và Ứng Dụng Thực Tiễn**”**. Qua đề tài, nhóm không chỉ muốn nắm rõ cách các kỹ thuật nhận dạng hoạt động, mà còn hiểu được quá trình phát triển và ứng dụng của chúng trong thực tế.

# CHƯƠNG 1 TỔNG QUAN

## **Giới thiệu đề tài**

Trong thế giới số hóa hiện nay, thị trường trí tuệ nhân tạo toàn cầu đang có nhiều bước tiến mạnh mẽ và với tốc độ đổi mới chưa từng có. Đặc biệt là lĩnh vực nhận dạng khuôn mặt, nó được áp dụng rộng rãi và đóng vai trò trong nhiều ứng dụng thực tiễn. Nhu cầu sử dụng các hệ thống nhận dạng khuôn mặt đang không ngừng tăng cao, khi mà việc xác thực nhanh, chính xác và không tiếp xúc trở nên quan trọng hơn bao giờ hết. Xuất phát từ nhu cầu thực tiễn và tiềm năng ứng dụng to lớn, thì nhóm đã nghiên cứu và lựa chọn đề tài **“**Tìm Hiểu Tổng Quan Về Các Kỹ Thuật Nhận Dạng Khuôn Mặt Và Ứng Dụng Thực Tiễn**”** nhằm tổng hợp, phân tích tất cả các kỹ thuật trong lĩnh vực nhận dạng khuôn mặt và triển khai thử nghiệm trên một số bài toán thực tế.

## **Nhiệm vụ của đề tài**

Nhiệm vụ của đề tài " Tìm Hiểu Tổng Quan Về Các Kỹ Thuật Nhận Dạng Khuôn Mặt Và Ứng Dụng Thực Tiễn " là tiến hành khảo sát, phân loại và phân tích các phương pháp nhận dạng khuôn mặt trong lĩnh vực thị giác máy tính. Đề tài tập trung làm rõ nguyên lý hoạt động, quy trình thực hiện, cũng như ưu nhược điểm của từng kỹ thuật. Từ đó, phân tích khả năng áp dụng của từng loại kỹ thuật vào từng bài toán, từng giai đoạn khác nhau của một dự án. Điều này giúp các nhà phát triển hiểu rõ và hoạch định trước quy trình phát triển của dự án. Và với các nhà đầu tư và doanh nghiệp, các dự án nhận dạng khuôn mặt có thể giúp họ dễ dàng quản lý nhân sự và học viên, nâng cao mức độ an toàn trong việc kiểm soát truy cập, xác minh danh tính trong giao dịch.

* + 1. Tính cấp thiết của đề tài

Trong bối cảnh hiện nay các nhà phát triển có thể học hỏi và khai thác cũng như mở rộng công nghệ một cách dễ tiếp cận hơn lúc trước, cùng với sự ra đời và cải tiến của nhiều công trình nghiên cứu mang tính lịch sử và có nhiều ông lớn trong lĩnh vực công nghệ tham gia sẽ tạo ra một môi trường cạnh tranh khốc liệt. Điều này đặt ra nhu cầu cấp bách cho các nhà nghiên cứu và doanh nghiệp công nghệ phải liên tục cập nhật, cải tiến và tối ưu hóa các thuật toán nhận dạng khuôn mặt nhằm nâng cao độ chính xác, tốc độ xử lý và khả năng ứng dụng thực tế.

Việc nghiên cứu tổng quan các kỹ thuật nhận dạng khuôn mặt là một bước quan trọng trong việc nâng cao hiệu quả của các hệ thống xác thực sinh trắc học. Việc này giúp các tổ chức và doanh nghiệp hiểu rõ đặc điểm, ưu, nhược điểm của từng phương pháp, từ đó lựa chọn giải pháp phù hợp với nhu cầu thực tế, qua đó giúp định hướng phát triển lâu dài với định hướng nhận dạng chính xác, thông minh và an toàn. Cụ thể hơn, việc tìm hiểu bao quát về các kỹ thuật nhận dạng có thể giúp các nhà phát triển và doanh nghiệp:

Phù hợp với bài toán thực tế: Mỗi kỹ thuật nhận dạng có ưu nhược điểm riêng. Việc hiểu rõ các phương pháp giúp lựa chọn công nghệ phù hợp với yêu cầu về độ chính xác, tốc độ và môi trường triển khai.

Tiết kiệm chi phí và tài nguyên: Việc lựa chọn đúng kỹ thuật giúp doanh nghiệp tối ưu hóa chi phí triển khai. Các mô hình hiệu quả sẽ giảm thiểu tài nguyên tính toán, đồng thời giảm thời gian xử lý, giúp doanh nghiệp tiết kiệm chi phí vận hành trong dài hạn.

Tăng khả năng mở rộng và linh hoạt trong ứng dụng: Các kỹ thuật nhận dạng khuôn mặt tiên tiến, khi được tích hợp vào các hệ thống, có thể mở rộng và áp dụng linh hoạt trong nhiều tình huống khác nhau.

Từ những lý do trên, có thể thấy rằng việc tìm hiểu tổng quan kỹ thuật nhận dạng không chỉ giúp doanh nghiệp nâng cao hiệu quả vận hành và bảo mật, mà còn thúc đẩy triển khai các giải pháp số thông minh, đáp ứng tốt hơn nhu cầu thị trường hiện đại.

* + 1. Ý nghĩa khoa học và thực tiễn của đề tài

Ý nghĩa khoa học: Đề tài đóng góp vào lĩnh vực thị giác máy tính và trí tuệ nhân tạo thông qua việc tổng hợp, phân tích và so sánh các kỹ thuật nhận dạng khuôn mặt hiện đại như Local Binary Pattern (LBP), Histogram of Oriented Gradients (HOG) và các mô hình deep learning tiên tiến khác. Việc khảo sát này không chỉ củng cố kiến thức nền tảng mà còn làm sáng tỏ phạm vi áp dụng của từng kỹ thuật trong các bài toán cụ thể như xác thực danh tính, điểm danh tự động, hay kiểm soát truy cập. Từ đó, cung cấp cơ sở khoa học vững chắc để các nhà phát triển và chuyên gia lựa chọn kỹ thuật phù hợp với mục tiêu và điều kiện triển khai thực tế.

Ý nghĩa thực tiễn: Đề tài hỗ trợ các doanh nghiệp trong việc lựa chọn và ứng dụng các kỹ thuật nhận dạng khuôn mặt phù hợp với nhu cầu thực tiễn. Việc áp dụng phù hợp không chỉ giúp nâng cao độ chính xác và tốc độ xử lý mà còn giảm thiểu chi phí vận hành, đồng thời tăng cường trải nghiệm người dùng và khả năng cạnh tranh trên thị trường. Ngoài ra, đề tài còn hỗ trợ các nhà phát triển và chuyên gia kỹ thuật trong việc nắm bắt xu hướng công nghệ mới, rút ngắn thời gian triển khai giải pháp thực tế, đồng thời cung cấp nền tảng dữ liệu và kiến thức có thể tái sử dụng hoặc mở rộng trong các hệ thống lớn hơn.

## **1.3 Mục tiêu nghiên cứu**

### 1.3.1 Mục tiêu tổng quát

Đề tài hướng đến việc nghiên cứu, tổng hợp và phân tích các kỹ thuật hiện đại trong lĩnh vực nhận dạng khuôn mặt, từ đó đánh giá khả năng ứng dụng của từng kỹ thuật trong các tình huống thực tế. Mục tiêu chính là đề xuất hướng triển khai phù hợp cho các doanh nghiệp có nhu cầu ứng dụng nhận diện khuôn mặt nhằm tối ưu hóa hoạt động vận hành, nâng cao hiệu quả quản lý và tăng cường trải nghiệm người dùng.

### 1.3.2 Mục tiêu cụ thể

Với bài nghiên cứu này, bước đầu tiên sẽ là thu thập và xử lý một tập dữ liệu hình ảnh khuôn mặt có gắn nhãn. Sau đó, nghiên cứu, so sánh và triển khai các kỹ thuật nhận dạng khuôn mặt như LBP, HOG và các phương pháp dựa trên kiến trúc CNN kết hợp để tìm ra phương pháp hiệu quả nhất. Quá trình nghiên cứu sẽ bao gồm các bước như phát hiện khuôn mặt, trích xuất đặc trưng, ánh xạ vector và so sánh độ tương đồng giữa các khuôn mặt. Mô hình sẽ được đánh giá dựa trên các tiêu chí như độ chính xác, tốc độ xử lý và khả năng hoạt động ổn định. Cuối cùng, đề tài sẽ đề xuất các giải pháp cụ thể và nền tảng cho các tổ chức và doanh nghiệp trong việc triển khai công nghệ nhận dạng khuôn mặt nhằm tối ưu hóa quy trình vận hành và nâng cao trải nghiệm người dùng.

**1.4 Đối tượng và phạm vi nghiên cứu**

### 1.4.1 Đối tượng nghiên cứu

Đối tượng nghiên cứu của đề tài là các kỹ thuật và phương pháp nhận dạng khuôn mặt, được sử dụng trong hệ thống xác thực người dùng. Qua các bước thu thập và xử lý dữ liệu ảnh tiếp đó thực hiện áp dụng tất cả các phương pháp trên một tập dữ liệu cụ thể. Qua đó đưa ra phương pháp tối ưu và đề xuất cải thiện và nâng cao chất lượng dự án.

1.4.2 Phạm vi nghiên cứu

Đề tài tập trung vào thu thập, xử lý dữ liệu ảnh khuôn mặt của người dùng từ tất cả nguồn dữ liệu có sẵn và đảm bảo tuân thủ các quy định về quyền riêng tư và bảo mật dữ liệu người dùng. Chúng tôi sẽ áp dụng các phương pháp phân tích dữ liệu và nhận dạng khuôn mặt để phân tích và so sánh hiệu quả của từng phương pháp, kĩ thuật cụ thể. Sau đó nghiên cứu đưa ra quy trình cải thiện và nâng cao chất lượng hệ thống, để đảm bảo rằng khi đưa vào thực tế thì dự án có thể mang đến độ chính xác, tốc độ và an toàn nhất. Điều này giúp mang lại giá trị lý thuyết và thực tiễn cho các doanh nghiệp trên thị trường trí tuệ nhân tạo

**1.5 Phương pháp nghiên cứu**

1.5.1 Phương pháp nghiên cứu sơ bộ

Trước khi tiến hành thu thập và xử lý dữ liệu, chúng tôi sẽ thực hiện một nghiên cứu sơ bộ để hiểu rõ hơn về nghiên cứu và các yếu tố quan trọng liên quan. Nghiên cứu này bao gồm việc tìm hiểu về các kỹ thuật nhận dạng khuôn mặt, ưu nhược điểm của từng phương pháp, và các yếu tố ảnh hưởng đến hiệu suất và độ chính xác của mô hình. Thông qua nghiên cứu sơ bộ, chúng tôi sẽ xác định các vấn đề cụ thể cần giải quyết và đề xuất các phương pháp nghiên cứu phù hợp.

1.5.2 Phương pháp nghiên cứu tài liệu

Chúng tôi sẽ tiến hành nghiên cứu tài liệu để thu thập thông tin về các phương pháp và công cụ nhận dạng khuôn mặt trong lĩnh vực thị giác máy tính và học máy. Qua việc đánh giá các nghiên cứu trước đây và các công trình khoa học liên quan, chúng tôi sẽ xác định các kỹ thuật nhận dạng khuôn mặt phù hợp nhất cho nghiên cứu của mình và áp dụng chúng vào việc triển khai các hệ thống xác thực người dùng.

1.5.3 Phương pháp nghiên cứu thống kê

Trong quá trình phân tích dữ liệu, chúng tôi sẽ sử dụng các phương pháp thống kê như phân tích phương sai hay kiểm tra độ tương quan hoặc sử dụng các phương pháp học máy (machine learning) để mô tả và phân tích các đặc trưng quan trọng trong nhận dạng khuôn mặt. Thông qua việc áp dụng các phương pháp thống kê này, chúng tôi sẽ đánh giá mối quan hệ giữa các đặc trưng và xác định các yếu tố ảnh hưởng đến độ chính xác và hiệu quả của hệ thống nhận dạng khuôn mặt.

1.5.4 Phương pháp thực nghiệm

Chúng tôi sẽ tiến hành thực nghiệm trên tập dữ liệu ảnh được thu thập từ các nguồn chính thống. Quá trình này bao gồm việc tiền xử lý dữ liệu hình ảnh, áp dụng các phương pháp nhận dạng khuôn mặt như LBP, HOG, hoặc các mô hình học sâu (CNN) để phân loại và xác minh khuôn mặt. Đồng thời, chúng tôi sẽ đánh giá hiệu quả của các kỹ thuật nhận dạng. Thông qua việc thực nghiệm này, chúng tôi sẽ kiểm tra và đảm bảo tính khả thi và hiệu quả của phương pháp nghiên cứu.

1.5.5 Phương pháp đánh giá

Cuối cùng, chúng tôi sẽ thực hiện phương pháp đánh giá để đo lường hiệu quả của các phương pháp nhận dạng khuôn mặt. Quá trình này bao gồm việc so sánh các chỉ số, độ chính xác, tốc độ xử lý, cũng như khả năng nhận dạng trong các điều kiện thực tế. Chúng tôi sẽ tiến hành đánh giá hiệu quả của từng kỹ thuật nhận dạng khuôn mặt trên từng bộ dữ liệu thực nghiệm, từ đó rút ra những nhận định về hiệu suất và khả năng áp dụng của mỗi phương pháp.

**1.6 Những đóng góp nghiên cứu của đề tài**

1.6.1 Đóng góp về mặt lý thuyết

Đề tài đóng góp vào việc áp dụng và phát triển các kỹ thuật nhận dạng khuôn mặt trong các hệ thống xác thực người dùng. Việc nghiên cứu và áp dụng các mô hình và thuật toán nhận dạng khuôn mặt, như các phương pháp học sâu (Deep Learning) và các mô hình phân loại, vào bài toán nhận diện các đặc trưng cùa khuôn mặc giúp mở rộng ứng dụng trong môi trường thực tế. Đề tài cũng cung cấp các giải pháp tối ưu hóa và cải thiện độ chính xác, từ đó tạo ra nhiều bước tiến vượt bậc cho các nghiên cứu sau này trong lĩnh vực nhận dạng khuôn mặt và bảo mật.

1.6.2 Đóng góp trong thực tiễn

Đề tài cung cấp các giải pháp và chiến lược cụ thể cho các doanh nghiệp trong việc áp dụng hệ thống xác thực người dùng bằng nhận dạng khuôn mặt. Việc cải thiện độ chính xác và hiệu quả của hệ thống xác thực giúp doanh nghiệp nâng cao trải nghiệm người dùng và tối ưu hóa quy trình xác thực. Đồng thời, các kết quả nghiên cứu có thể giúp các doanh nghiệp xây dựng các hệ thống bảo mật mạnh mẽ hơn, giảm thiểu rủi ro gian lận, và tăng cường sự tin cậy từ phía khách hàng. Điều này có thể dẫn đến việc nâng cao chất lượng dịch vụ và tăng trưởng trong môi trường thương mại điện tử ngày càng phát triển.

# **CHƯƠNG 2 CƠ SỞ LÝ THUYẾT**

## **2.1** **Tổng quan về nhận dạng khuôn mặt**

### 2.1.1 Giới thiệu về nhận dạng khuôn mặt (Face Recogniton)

Nhận diện khuôn mặt thực chất là quá trình xác định danh tính dựa vào đặc điểm sinh trắc học của khuôn mặt thông qua hình ảnh hoặc video. Công nghệ này kết hợp các lĩnh vực như thị giác máy tính và học sâu để phân tích và đánh giá đặc trưng khuôn mặt.

* Phát hiện khuôn mặt: Xác định vị trí có khuôn mặt của người trong ảnh hay video.
* Căn chỉnh khuôn mặt: Đưa khuôn mặt về chính giữa, xác định các điểm trên khuôn mặt như: mắt, mũi, miệng, sau đó xoay hay cắt ảnh sao cho các điểm trở nên cân đối.
* Trích xuất các đặc trưng: Hình ảnh sau căn chỉnh được chuyển thành chuỗi vector số học chứa các đặc trưng nhận diện, nói đơn giản là chuyển thông tin khuôn mặt thành dữ liệu máy tính có thể xử lý.
* So sánh với các dữ liệu đã lưu: Vector đặc trưng này sẽ được đối chiếu với các vector đã lưu trong hệ thống, thuật toán tự động nhận diện sự tương đồng để xác định danh tính nếu có khớp.
* Kết quả output: Với trường hợp xác thực thành công, hệ thống sẽ trả về định danh người dùng tương ứng. Nếu không có sự trùng khớp, hệ thống sẽ từ chối truy cập.

### 2.1.2 Ưu điểm và hạn chế

Ưu điểm:

* Tự động hóa cao: Giúp con người hay các doanh nghiệp có thể giảm bớt đi khối lượng công việc, thay thế trong các việc lặp đi lặp lại.
* Tốc độ xử lý nhanh: Có thể phân tích các video và hình ảnh trong thời gian ngắn để đưa ra các kết quả chính xác.
* Độ chính xác cao: Khi được huấn luyện tốt thì mô hình có thể đưa ra được kết quả chính xác cao, ổn định.
* Ứng dụng rộng rãi: Có thể áp dụng trong nhiều lĩnh vực như: nhận diện, y tế, an ninh, công nghiệp, ô tô tự lái, nông nghiệp, …

Hạn chế:

* Phụ thuộc dữ liệu: Cần lượng dữ liệu lớn và chất lượng để huấn luyện cho mô hình.
* Yêu cầu phần cứng mạnh: Các mô hình Deep Learning cần GPU để hoạt động hiệu quả.
* Ứng dụng nhận diện khuôn mặt có thể ảnh đến quyền riêng tư cá nhân và cũng có thể dùng trong việc lừa đảo.

## **2.2 Thư viện OpenCV (Open Source Computer Vision Library)**

OpenCV là một thư viện mã nguồn mở cho thị giác máy tính, học máy và xử lý hình ảnh. Đóng một vai trò quan trọng trong hoạt động thời gian thực. Bằng cách đó, người ta có thể xử lý hình ảnh và video để xác định các đối tượng, khuôn mặt hoặc thậm chí là nhận diện.

Khi nó được tích hợp với các thư viện khác nhau, chẳng hạn như NumPy, Python có khả năng xử lý cấu trúc mảng opencv để phân tích. Để xác định một mẫu hình ảnh và các tính năng khác nhau của nó, sử dụng không gian vector và thực hiện các phép toán trên các tính năng này.

OpenCV cho phép thực hiện các thao tác khác nhau trong hình ảnh như:

* Đọc hình ảnh : OpenCV giúp đọc hình ảnh từ tệp hoặc trực tiếp từ máy ảnh để có thể truy cập và thực hiện các bước xử lý thêm.
* Tăng cường hình ảnh: Có thể nâng cao hình ảnh bằng cách điều chỉnh độ sáng, độ sắc nét hoặc độ co lại của hình ảnh..
* Phát hiện đối tượng: Đối tượng cũng có thể được phát hiện bằng cách sử dụng OpenCV,. Điều này cũng có thể nhận dạng khuôn mặt, hình dạng hoặc thậm chí là đối tượng.
* Lọc hình ảnh: Thay đổi hình ảnh bằng cách áp dụng các bộ lọc khác nhau như làm mờ hoặc làm sắc nét.
* Nhận dạng đối tượng
* Học máy và phân cụm (ml, flann)
* Tăng tốc CUDA (gpu)
  1. **Các phương pháp, kỹ thuật tiền xử lý ảnh.**
     1. Làm mịn ảnh (Image Smoothing-blurring)

Trong quá trình tiền xử lý và đầu vào nhận diện, ảnh thường chứa nhiều nhiễu hoặc chi tiết nhỏ không cần thiết gây khó khăn cho việc trích xuất đặc trưng. Để giải quyết vấn đề này, kỹ thuật làm mịn ảnh hay còn gọi là làm mờ được áp dụng giúp giảm mức độ biến thiên giữa các pixel lân cận, giúp ảnh trở nên mượt hơn và loại bỏ nhiễu cục bộ.

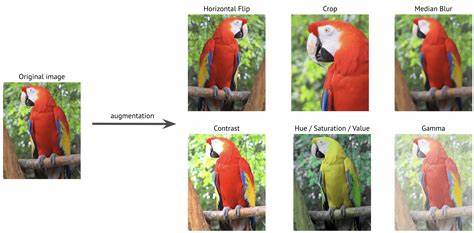
* Làm mờ trung bình: Mỗi pixel được thay bằng trung bình cộng của các pixel xung quanh, sử dụng khi nhiễu không đồng nhất và ít tập trung.
* Làm mờ Gaussian: Sử dụng phân phối chuẩn để tính trung bình có trọng số, sử dụng khi cần làm mờ mịn và giảm nhiễu hiệu quả mà vẫn giữ biên rõ ràng
* Làm mờ trung vị: Mỗi pixel được thay bằng giá trị trung vị của vùng xung quanh, sử dụng khi ảnh bị nhiễu muối tiêu (một số pixel trong ảnh bị thay đổi đột ngột thành màu đen hoàn toàn (0) hoặc trắng hoàn toàn (255))

### 2.3.2 Phóng đại dữ liệu (Data Augmentation)

Trong nhận dạng khuôn mặt, dữ liệu đầu vào thường bị giới hạn về số lượng hoặc thiếu tính đa dạng. Khiến mô hình dễ bị overfitting. Kỹ thuật Data được sử dụng nhằm khắc phục vấn đề này bằng cách tạo ra nhiều phiên bản biến đổi khác nhau của ảnh gốc nhưng vẫn giữ nguyên nhãn hoặc nội dung chính, từ đó giúp mô hình học được tốt hơn và tổng quát hơn.

Các kỹ thuật phóng đại dữ liệu phổ biến gồm:

* Lật ảnh : Lật ngang thường được sử dụng trong nhận diện khuôn mặt vì khuôn mặt thường đối xứng.
* Xoay ảnh: Xoay ảnh với một góc nhất định như ±15° để mô phỏng các tư thế đầu khác nhau.
* Phóng to/thu nhỏ: Làm thay đổi tỷ lệ kích thước đối tượng trong ảnh.
* Dịch chuyển :Di chuyển ảnh theo trục X hoặc Y.
* Thay đổi độ sáng, độ tương phản: Mô phỏng điều kiện ánh sáng khác nhau.
* Thêm nhiễu : Làm cho dữ liệu giống với môi trường thực hơn. Giúp mô hình thích ứng tốt với ảnh thực tế bị nhiễu vd: camera chất lượng thấp.
* Cắt ảnh ngẫu nhiên :cắt các vùng nhỏ ngẫu nhiên từ ảnh gốc để tạo thêm các biến thể.



<https://th.bing.com/th/id/OIP.Za3VLUEHu7JhiLRWO3Lv_AHaDp?cb=iwc2&rs=1&pid=ImgDetMain>

## **Phát hiện đặc trưng và mô tả (Feature Detection & Description)**

Phát hiện đặc trưng là bước quan trọng nhằm tìm ra các điểm đặc trưng như khóe mắt, sống mũi, miệng, cằm... giúp mô hình có thể phân biệt giữa các khuôn mặt khác nhau. Từ đó phục vụ cho các tác vụ so khớp, xác thực hoặc phân loại khuôn mặt.

### 2.4.1 Phát hiện khuôn mặt bằng MTCNN

MTCNN (Multi-task Cascaded Convolutional Networks) là một phương pháp phát hiện khuôn mặt hiện đại, tích hợp hai tác vụ chính là phát hiện khung khuôn mặt và dự đoán điểm đặc trưng trên khuôn mặt như: mắt, mũi, miệng,...

Mô hình này hoạt động theo cấu trúc gồm ba mạng nơ-ron tích chồng:

* P-Net (Proposal Network): thực hiện quét sơ bộ ảnh để đề xuất các vùng có thể chứa khuôn mặt.
* R-Net (Refine Network): lọc lại các vùng được P-Net đề xuất và loại bỏ các hộp dự đoán sai.
* O-Net (Output Network): cung cấp vị trí chính xác hơn của khuôn mặt và 5 điểm đặc trưng (2 mắt, mũi, 2 khóe miệng) để phục vụ căn chỉnh khuôn mặt.

#### Trích xuất đặc trưng khuôn mặt bằng FaceNet

FaceNet là một mô hình học sâu nhằm ánh xạ mỗi ảnh khuôn mặt vào một không gian đặc trưng có kích thước cố định (128 chiều hoặc 512 chiều), gọi là face embedding. Khoảng cách giữa hai vector embedding phản ánh mức độ giống nhau giữa hai khuôn mặt.

Đặc điểm của embedding từ FaceNet:

* Hai khuôn mặt giống nhau sẽ có embedding gần nhau (khoảng cách nhỏ).
* Hai khuôn mặt khác nhau sẽ có khoảng cách embedding lớn hơn.

Để xác định xem 2 khuôn mặt có giống nhau không thì ta cần so sánh độ gần nhau giữa hai vector embedding.

Ta sử dụng khoảng cách Euclidean là độ dài đoạn thẳng giữa hai vector trong không gian nhiều chiều.

Công thức khoảng cách Euclidean:

A black square with numbers and a square in it

AI-generated content may be incorrect.

Nếu hai khoảng cách hai vector gần nhau trong không gian, thì hai khuôn mặt tương ứng giống nhau

* 1. **Mô Hình**
     1. Logistic Regression

**2.5.1.1 Giới thiệu về mô hình Logistic Regression**

Logistic Regression là một mô hình học máy tuyến tính thường được sử dụng trong các bài toán phân loại nhị phân (2 lớp). Khác với các mạng neural sâu như MobileNet, Logistic Regression có cấu trúc đơn giản hơn, phù hợp cho các hệ thống cần giải thích rõ ràng và có dữ liệu giới hạn. Dù không chuyên dụng cho nhận diện khuôn mặt, Logistic Regression có thể đóng vai trò như một lớp phân loại đầu ra khi xử lý đặc trưng từ các mô hình phức tạp hơn.

**2.5.1.2 Nền tảng toán học**

Logistic Regression sử dụng hàm sigmoid để ánh xạ giá trị đầu ra của một tổ hợp tuyến tính vào khoảng (0, 1), giúp xác suất hóa dự đoán.

Biểu thức cơ bản:

z = w^T x + b

\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}}

Trong đó:

* x: vector đầu vào
* w: vector trọng số
* b: hệ số chệch (bias)
* \hat{y}: xác suất đầu ra (phân loại)
* \sigma(z): hàm sigmoid

Hàm mất mát được sử dụng phổ biến là Binary Cross-Entropy:

L = -[y \log(\hat{y}) + (1 - y)\log(1 - \hat{y})]

**2.5.1.3 Diễn giải thuật toán**

Giai đoạn 1:

* Dữ liệu được tiền xử lý và chuyển thành vector đầu vào x.
* Tính tổ hợp tuyến tính z = w^T x + b.

Giai đoạn 2:

* Sử dụng hàm sigmoid để chuyển giá trị z thành xác suất thuộc một trong hai lớp.
* Nếu cần, áp dụng ngưỡng (threshold) để phân loại.

**2.5.1.4 Phân tích độ phức tạp**

Do Logistic Regression chỉ sử dụng tổ hợp tuyến tính và hàm sigmoid, số phép tính rất thấp. Độ phức tạp tính toán tuyến tính theo số đặc trưng (features), phù hợp với các thiết bị giới hạn tài nguyên và các bài toán phân tích nhanh.

**2.5.1.5 Ưu Nhược điểm**

**Ưu điểm:**

* Dễ hiểu, dễ triển khai.
* Thời gian huấn luyện nhanh.
* Không yêu cầu tài nguyên tính toán cao.
* Dễ giải thích kết quả, phù hợp với các ứng dụng yêu cầu giải thích (interpretability).

**Nhược điểm:**

* Không xử lý tốt các bài toán phi tuyến tính phức tạp.
* Hiệu quả kém hơn khi dữ liệu có mối quan hệ phi tuyến mạnh.
* Dễ bị ảnh hưởng bởi nhiễu hoặc các đặc trưng không liên quan.
  + 1. SVM

### 2.5.5 Support Vector Machine (SVM)

#### 2.5.5.1 Giới thiệu về mô hình SVM

Support Vector Machine (SVM) là một mô hình học máy mạnh mẽ được sử dụng cho các bài toán phân loại (classification) và hồi quy (regression). Đặc biệt hiệu quả trong các bài toán phân loại nhị phân với biên phân cách rõ ràng. SVM có thể hoạt động tốt cả trong không gian tuyến tính lẫn phi tuyến thông qua kỹ thuật kernel trick, biến nó trở thành lựa chọn phổ biến trong nhận diện khuôn mặt, phân loại văn bản, và phát hiện bất thường.

#### 2.5.5.2 Nền tảng toán học

SVM tìm một siêu phẳng (hyperplane) tối ưu phân tách các điểm dữ liệu thành hai lớp sao cho khoảng cách từ siêu phẳng đến các điểm gần nhất của hai lớp (support vectors) là lớn nhất.

Bài toán tối ưu (cho SVM tuyến tính):

\min\_{w,b} \frac{1}{2} \|w\|^2 \quad \text{với ràng buộc } y\_i(w^T x\_i + b) \geq 1

Trong đó:

* w: vector trọng số
* b: hệ số chệch
* x\_i: điểm dữ liệu
* y\_i \in \{-1, +1\}: nhãn của dữ liệu

Để giải quyết các bài toán phi tuyến, ta dùng kernel function K(x\_i, x\_j) để ánh xạ dữ liệu vào không gian có thể phân tách tuyến tính:

* Kernel phổ biến: Linear, Polynomial, RBF (Radial Basis Function), Sigmoid

#### 2.5.5.3 Diễn giải thuật toán

Giai đoạn 1:

* Chuyển dữ liệu về dạng chuẩn (standardization) để tối ưu hiệu quả phân tách.
* Xây dựng hàm mất mát SVM và chọn hàm kernel phù hợp.

Giai đoạn 2:

* Huấn luyện mô hình để tìm được siêu phẳng tối ưu hoặc ranh giới phân lớp trong không gian kernel.
* Mô hình được xác định chủ yếu bởi các điểm hỗ trợ (support vectors) – những điểm nằm gần ranh giới phân cách.

#### 2.5.5.4 Phân tích độ phức tạp

Độ phức tạp phụ thuộc vào số lượng mẫu dữ liệu và số chiều của đặc trưng:

* Huấn luyện có độ phức tạp từ O(n^2) đến O(n^3), với n là số mẫu.
* Với dữ liệu lớn, cần dùng SVM tuyến tính hoặc các thuật toán tối ưu hóa như SMO để tăng hiệu suất.
* Khi số mẫu nhỏ hơn số đặc trưng (n < d), SVM thường cho kết quả tốt.

#### 2.5.5.5 Ưu Nhược điểm

**Ưu điểm:**

* Chính xác cao, đặc biệt trong bài toán phân lớp rõ ràng.
* Hiệu quả với dữ liệu có số chiều lớn.
* Có thể xử lý phi tuyến với kernel.
* Khả năng tổng quát hóa tốt, ít bị overfitting.

**Nhược điểm:**

* Không hiệu quả với dữ liệu rất lớn hoặc có nhiều nhiễu.
* Thời gian huấn luyện cao với dữ liệu lớn.
* Khó lựa chọn và điều chỉnh kernel phù hợp.
* Kết quả khó giải thích hơn so với Logistic Regression.
  + 1. Random Forest

……………………………………….

* + 1. FaceNet

…………………………………………………………………………….

2.5.4 MobileNetV2

## *2.5.3.1 Giới thiệu về mô hình MobileNet*

MobileNet là một dòng mạng neural sâu, nó không phải là một mô hình nhận diện khuôn mặt chuyên biệt như FaceNet, nhưng nó thường được dùng làm kiến trúc nền (backbone) [ct] trong các hệ thống xác thực khuôn mặt để tối ưu hóa cho các thiết bị có tài nguyên hạn chế. Mục tiêu chính của MobileNet là duy trì hiệu suất tốt trong khi giảm thiểu số lượng tham số và chi phí tính toán.

*2.5.3.2 Nền tảng toán học*

MobileNet thực hiện các phép tính theo kiểu đơn giản, chia nhỏ công việc thay vì dùng một phép tích chập lớn như các mạng CNN truyền thống.

Bước 1: Mở rộng chiều dữ liệu (Expansion Layer) [ct]: ta dùng một phép tính convolution [ct] 1x1 để nhân bản thông tin sau đó dữ liệu sẽ trở nên dày hơn, giúp các bước sau xử lý tốt hơn.

*z = ReLU6(W1​∗x)* (2.)

Trong đó:  
- x: đầu vào  
- W1​: bộ lọc 1×1 dùng để tăng số kênh  
- z: đầu ra sau bước mở rộng

Và được giới hạn bởi hàm kích hoạt ReLU [ct]

*ReLU6(x)=min(max(0,x),6)*

Diễn giải:

* Nếu x ≤ 0 : đầu ra là 0
* Nếu 0 < x <6 : đầu ra là chính x
* Nếu x ≥ 6 : đầu ra bị giới hạn ở mức 6

Dùng ReLU6 vì giá trị 6 vừa đủ lớn để giữ tính phi tuyến và vừa đủ nhỏ để dễ lượng tử hóa xuống 8 – bit, đây là giá trị tối ưu để mô hình chạy nhẹ và nhanh trên thiết bị di dộng.

Bước 2: Tích chập Depthwise [ct], thay vì ta dùng một phép tích chập toàn bộ thì Depthwise thực hiện tích chập độc lập mỗi kênh để giảm mạnh số phép tính và tham số.

*di​=Ki​∗ zi​*

Trong đó:  
- Ki ​: bộ lọc áp dụng riêng cho kênh i  
- zi: kênh thứ i của đầu ra từ bước 1  
- di​: kết quả tích chập depthwise kênh i

Bước 3: Projection Layer [ct]: Sau khi các phép tích chập hoàn thành công việc sẽ được kết hợp các kênh với convolution 1x1 để đưa về số kênh ban đầu.

*y=W2​∗ d (2. )*

Trong đó:  
- W2​: bộ lọc 1×1 giảm số kênh  
- y: đầu ra cuối cùng của block

Bước 4: Residual Connection [ct]: Nếu đầu vào và đầu ra có cùng kích thước thì ta cộng lại đầu vào với kết quả đầu ra, giúp giữ lại thông tin gốc, tránh mất mát, và giúp máy học nhanh hơn.

*x* + *y (2. )*

*2.5.3.3 Diễn giải thuật toán*

Giai đoạn 1:

* Ảnh được đưa vào MobileNetV2, qua chuỗi các khối Inverted Residual [ct] và Linear Bottleneck [ct].
* Các lớp depthwise và convolution giúp giảm tính toán nhưng vẫn giữ được đặc trưng quan trọng.

Giai đoạn 2:

* Với nhận diện khuôn mặt, đầu ra là vector đặc trưng (embedding).
* Vector này có thể:
  + Dùng để huấn luyện bộ phân loại.
  + So sánh với vector từ các ảnh khác (ví dụ: nhận diện người lạ hay người quen).

*2.5.3.4 Phân tích độ phức tạp*

MobileNetV2 sử dụng một kỹ thuật tích chập đặc biệt giúp chia nhỏ quá trình xử lý thành hai bước đơn giản hơn thay vì làm tất cả cùng lúc. Đồng thời, còn sử dụng các khối xử lý giúp giữ lại thông tin quan trọng mà vẫn tiết kiệm được công sức tính toán. Nhờ đó, tổng số phép tính của mạng giảm nhiều lần so với các mô hình học sâu thông thường.

*2.5.3.5 Ưu Nhược điểm*

Ưu điểm:

* Nhẹ, nhanh, dễ triển khai trên thiết bị có tài nguyên hạn chế.
* Hiệu quả tốt với tốc độ xử lý nhanh.
* Phù hợp làm backbone cho nhiều ứng dụng như nhận diện khuôn mặt, phân loại ảnh, v.v.

Nhược điểm:

* Độ chính xác thấp hơn so với các mô hình lớn hơn như ResNet50 hay EfficientNet.
* Có thể bỏ sót các chi tiết nhỏ trong ảnh vì tối ưu chủ yếu cho tốc độ.
  1. **Nhận dạng đối tượng (Object Recognition)**

Nhận dạng đối tượng là quá trình xác định sự hiện diện và loại của một đối tượng cụ thể trong ảnh, điều này đồng nghĩa với việc không chỉ phát hiện có khuôn mặt, mà còn biết đó là ai trong danh sách đã được ghi nhớ.

* Thu nhận hình ảnh: Camera liên tục ghi lại các khung hình video làm dữ liệu đầu vào.
* Xử lý hình ảnh: Khung hình được xử lý giảm nhiễu, điều chỉnh màu và chuẩn hóa định dạng.
* Phát hiện khuôn mặt: Hệ thống xác định vị trí khuôn mặt và tách riêng vùng chứa khuôn mặt.
* Trích xuất đặc trưng: Mỗi khuôn mặt được chuyển thành một vector đặc trưng duy nhất.
* Nhận diện danh tính: Vector được đưa vào mô hình phân loại để so sánh và xác định danh tính.
* Hiển thị kết quả: Hệ thống đánh giá độ tin cậy và hiển thị tên cùng tỉ lệ chính xác qua độ tin cậy trên video.